

RECEIVED

NOV 26 2004

Technology Center 2600 Attorney Docket No. JP919990202



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): Y. Tahara et al.
Docket No.: JP919990202
Serial No.: 09/656,964
Filing Date: September 7, 2000
Group: 2655
Examiner: Michael N. Opsasnick

I hereby certify that this paper is being deposited on this date with the U.S. Postal Service as first class mail addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Signature: Lisa L. Vulpis Date: November 19, 2004

Title: Methods and Apparatus for Recognized Word
Registration in Accordance With Speech
Recognition

THIRD SUPPLEMENTAL APPEAL BRIEF

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Applicants (hereinafter referred to as "Appellants") hereby appeal the rejection of claims 1-3, 5-7 and 9-11 of the above-identified application.

REAL PARTY IN INTEREST

The present application is assigned to International Business Machines Corporation, as evidenced by an assignment recorded December 4, 2000 in the U.S. Patent and Trademark Office at Reel 011310, Frame 0622. The assignee, International Business Machines Corporation, is the real party in interest.

RELATED APPEALS AND INTERFERENCES

There are no known related appeals and interferences.

STATUS OF CLAIMS

Claims 1-3, 5-7 and 9-11 stand rejected under 35 U.S.C. §103(a). Claims 1-3, 5-7 and 9-11 are appealed. Claims 4, 8 and 12 have been allowed.

STATUS OF AMENDMENTS

There has been no amendment filed subsequent to the final rejection. However, a Response to Final Office Action was filed on February 27, 2003, an Appeal Brief was filed on June 5, 2003, a Supplemental Appeal Brief was filed on November 26, 2003, and a Second Supplemental Appeal Brief was filed on May 25, 2004.

SUMMARY OF INVENTION

The present invention provides techniques that maintain speech recognition accuracy when a new word is being registered in a speech recognition dictionary (Specification, page 3, lines 1-3).

In one aspect of the invention, a technique for performing recognized word registration may comprise the following steps/operations. First, a word inscription specified by a user is obtained. A word dictionary is searched to obtain a sounds-like spelling corresponding to the word inscription. A pronunciation dictionary is searched to obtain a base form corresponding to the sounds-like spelling that has been obtained. Then, the base form is registered in a speech recognition dictionary (Specification, page 3, line 23, through page 4, line 3).

In another aspect of the invention, techniques for performing recognized word registration may comprise the following steps/operations. First, a word inscription is specified by a user. A word dictionary is searched to obtain a plurality of sounds-like spellings that correspond to said word inscription and sounds-like spelling scores that correspond to the sounds-like spellings. The plurality of sounds-like spellings are displayed for the user. The sounds-like spelling that is selected by the user is obtained from among the plurality of sounds-like spellings. Then, a pronunciation dictionary is searched to obtain a base form and a pronunciation score corresponding to the sounds-like spelling that has been obtained. A determination is made as to whether the pronunciation score exceeds a predetermined threshold value. The base form is then registered in a speech recognition dictionary

when the pronunciation score exceeds the predetermined threshold value (Specification, page 4, lines 4-19).

In a further aspect of the invention, techniques for performing recognized word registration may comprise the following steps/operations. First, it is determined whether first voice information obtained for a user's voice matches a predetermined condition. A speech recognition wizard panel that includes a new word input field and a sounds-like spelling input field is displayed on a display screen, when the voice information matches said predetermined condition. A new word and a sounds-like spelling that are entered in the speech recognition wizard panel are obtained. Second voice information based on the user's pronunciation provided for the new word and the sounds-like spelling is obtained. The second voice information, the new word and the sounds-like spelling are employed to specifically describe a base form. Then, the base form is added to a speech recognition dictionary (Specification, page 4, line 20, through page 5, line 7).

Accordingly, by way of example, the following steps describe an illustrative embodiment of the present invention. A sounds-like spelling group is generated in which scores are used for word descriptions. A user selects a correct sounds-like spelling from a generated sounds-like spelling group, and a group of base forms is generated with scores provided for the selected sounds-like spelling and the reading of the word. In this fashion, a base form having a score that exceeds a reference value is registered without a voice recording being required. During the speech recognition process as performed by a user, when a predetermined error state is detected, e.g., when the number of recognition errors exceeds N, a panel is output requesting that the user register the inscription, the sounds-like spelling and the pronunciation inscription for the word and that the user record a corresponding pronunciation. In accordance with the pronunciation provided by the user, the base form is obtained and is re-registered in the speech recognition dictionary (Specification, page 3, lines 13-22).

FIGs. 1 and 2 conceptually depict recognized word registration processing techniques, according to an embodiment of the invention (Specification, page 13, lines 13-26). FIGs. 5 and 6 respectively depict a word dictionary and a pronunciation dictionary, according to embodiments of the invention (Specification, page 16, line 14, through page 17, line 5). An embodiment of a

recognized word registration methodology of the invention is shown in FIG. 7, with reference to user interfaces in FIGs. 8, 9 and 10 (Specification, page 17, line 11, through page 19, line 9). An embodiment of a speech recognition methodology of the invention is shown in FIG. 11, with reference to a user interface in FIG. 12 (Specification, page 19, line 11, through page 20, line 15).

ISSUE PRESENTED FOR REVIEW

Whether claims 1-3, 5-7 and 9-11 are properly rejected under 35 U.S.C. §103(a) as being unpatentable over U.S. Patent No. 5,208,897 to Hutchins (hereinafter "Hutchins"), in view of U.S. Patent No. 6,233,553 to Contolini et al. (hereinafter "Contolini") and IBM Technical Disclosure Bulletin, Vol. 35, Issue 1a, p. 59 (hereinafter "IBM TDB").

GROUPING OF CLAIMS

Claims 1, 5 and 9 stand or fall together. Claims 2, 6 and 10 stand or fall together. Claims 3, 7 and 11 stand or fall together.

ARGUMENT

Appellants incorporate by reference herein the disclosure of all previous responses filed in the present application, namely: an Amendment and Response to Office Action dated September 18, 2002; a Response to Final Office Action dated February 27, 2003; an Appeal Brief dated June 5, 2003; a Supplemental Appeal Brief dated November 26, 2003, and a Second Supplemental Appeal Brief dated May 25, 2004.

With regard to the issue of whether claims 1-3, 5-7 and 9-11 are unpatentable under 35 U.S.C. §103(a) over Hutchins in view of Contolini and IBM TDB, the Office Action contends that the cited combination discloses all of the claim limitations recited in the subject claims. Appellants respectfully assert that the combination of Hutchins, Contolini and IBM TDB fails to establish a prima facie case of obviousness under 35 U.S.C. §103(a), as specified in M.P.E.P. §2143.

As set forth in M.P.E.P. §2143, three requirements must be met to establish a prima facie case of obviousness. First, there must be some suggestion or motivation to combine reference teachings. Second, there must be a reasonable expectation of success. Third, the cited combination must teach or suggest all the claim limitations. While it is sufficient to show that a prima facie case of obviousness has not been established by showing that one of the requirements has not been met, Appellants respectfully believe that none of the requirements have been met.

First, there is a clear lack of motivation to combine the references. For at least this reason, a prima facie case of obviousness has not been established. Hutchins is directed to the performance of speech recognition, while Contolini is directed to the generation of a phonetic transcription associated with a spelled word and IBM TDB is directed to building word models using sounds-like spellings. That is, the teachings in each reference are directed to completely different processes in speech recognition technology; one (Hutchins) toward actual real-time recognition of a spoken utterance, the other two (Contolini and IBM TDB) toward building models that may eventually be used in actual real-time recognition of a spoken utterance. However, other than a very general and conclusory statement in the Office Action, there is nothing in the three references that reasonably suggests why one would actually combine the teachings of these three references.

The Federal Circuit has stated that when patentability turns on the question of obviousness, the obviousness determination “must be based on objective evidence of record” and that “this precedent has been reinforced in myriad decisions, and cannot be dispensed with.” In re Lee, 277 F.3d 1338, 1343 (Fed. Cir. 2002). Moreover, the Federal Circuit has stated that “conclusory statements” by an examiner fail to adequately address the factual question of motivation, which is material to patentability and cannot be resolved “on subjective belief and unknown authority.” Id. at 1343-1344.

In the Office Action at page 3, paragraph 2, the Examiner provides the following statement to prove motivation to combine Hutchins and Contolini, with emphasis supplied: “[t]herefore, it would have been obvious . . . to modify the teachings of Hutchins (5208897) with a trained sound dictionary of words as specified by the user because it would advantageously allow for a higher recognition rate tailored for individual users and more efficient dictionaries.”

Additionally, in the Office Action at page 4, paragraph 1, the Examiner provides the following statement to prove motivation to combine Hutchins and IBM TDB, with emphasis supplied: “[t]herefore, it would have been obvious . . . to modify the teachings of Hutchins in view of Contolini (6233553) with a ‘sounds like spelling’ technique because it would advantageously allow user to enter the information more accurately than the phonetic pronunciations.”

Since Hutchins has no disclosure relating to word registration, it is unclear why there would be any motivation to combine Hutchins with Contolini or IBM TDB. Although the Examiner cites “IBM TDB, disclosure text, near the end,” after the above statement, Appellants submit that this statement is based on the type of “subjective belief and unknown authority” that the Federal Circuit has indicated provides insufficient support for an obviousness rejection. More specifically, the Examiner fails to identify any objective evidence of record which supports the proposed combination.

Second, Appellants assert that there is no reasonable expectation of success in achieving the present invention through a combination of Hutchins, Contolini and IBM TDB. For at least this reason, a prima facie case of obviousness has not been established. Despite the assertion in the Office Action, Appellants do not believe that Hutchins, Contolini and IBM TDB are combinable since it is not clear how one would combine them. There is no guidance provided in the present Office Action. However, even if combined, for the sake of argument, they would not achieve the automated registration techniques of the claimed invention.

Third, Appellants assert that even if combined, the combination fails to teach or suggest all of the limitations of the claims. For at least this reason, a prima facie case of obviousness has not been established.

Regarding independent claims 1, 5 and 9, the invention recites a method, apparatus and program code, respectively, for performing recognized word registration. First, a word inscription specified by a user is obtained. A word dictionary is searched to obtain a sounds-like spelling corresponding to the word inscription. A pronunciation dictionary is searched to obtain a base form corresponding to the sounds-like spelling that has been obtained. Then, the base form is registered in a speech recognition dictionary.

In the Office Action, Hutchins is cited as teaching such steps with the exception of the concept of continued adding to a dictionary and use of sounds-like spellings, which is said to be taught by Contolini and IBM TDB, respectively. Appellants strongly disagree. As explained above, all of Hutchins' operations are associated with the actual recognition of speech uttered by a user. The techniques of Hutchins do not relate to word registration in a speech recognition dictionary, as recited in the elements of claims 1, 5 and 9.

For example, in rejecting the first element of claims 1, 5 and 9, reciting the obtaining of a word inscription specified by a user, the Examiner simply directs the Appellants to FIG. 4a of Hutchins. FIG. 4a discloses audio input in a word recognition technique and fails to disclose a word inscription specified by a user in a word registration method.

In rejecting the second element of claims 1, 5 and 9, reciting the searching of a word dictionary to obtain a sounds-like spelling corresponding to said word inscription, the Examiner directs the Appellants to column 9, line 48 through column 11, line 42 of Hutchins. The Examiner has acknowledged that a word dictionary having sounds-like spellings are not disclosed in Hutchins, but contends that the remainder of this claim element is disclosed in these sections. However, these sections disclose evaluations of subsyllables and syllables from incoming speech against a predetermined database of subsyllable and syllable spellings. Simply accessing a database for syllable spellings, as described in Hutchins, does not provide proper disclosure for the searching of a word dictionary to obtain a (sounds-like) spelling corresponding to a word inscription input by a user in a word registration method.

In rejecting the third element of claims 1, 5 and 9, reciting the searching of a pronunciation dictionary to obtain a base form corresponding to the sounds-like spelling that has been obtained, the Examiner directs the Appellants to a portion of Hutchins that discloses phonetic detection in a digitizer of audio input. It is not clear how this relates to this third element. There is no disclosure of a base form of a word or of the searching of a dictionary to obtain a base form of a word based on a sounds-like spelling. Therefore, Hutchins also fails to disclose this element.

Finally, in rejecting the fourth element of claims 1, 5 and 9, reciting the registering of the base form in a speech recognition dictionary, the Examiner again directs the Appellants to the section

of a base form of a word or of the searching of a dictionary to obtain a base form of a word based on a sounds-like spelling. Therefore, Hutchins also fails to disclose this element.

Finally, in rejecting the fourth element of claims 1, 5 and 9, reciting the registering of the base form in a speech recognition dictionary, the Examiner again directs the Appellants to the section of Hutchins relating to evaluations of subsyllables and syllables. However, Hutchins fails to disclose registering a base form of a word in a speech recognition dictionary in a word recognition method.

The Examiner acknowledges that Hutchins teaches away from adding to a dictionary since Hutchins does not relate to a word registration process. However, the Examiner cites Contolini in an attempt to incorporate continued adding to a dictionary. Contolini discloses an automatic determination of phonetic transcriptions associated with spelled words. A tool allows a user to specify a confidence level, and a system stores all generated pronunciations that fulfill the confidence level, updating a pronunciation dictionary or lexicon.

The combination of Hutchins and Contolini fails to remedy all of the deficiencies of Hutchins presented above. For example, the combination fails to disclose the searching of a dictionary to obtain a sounds-like spelling corresponding to a word inscription. As admitted by the Examiner, Contolini contains no disclosure of sounds-like spellings. The combination also fails to disclose the searching of a pronunciation dictionary to obtain a base form corresponding to a sounds-like spelling. Contolini only discloses the conversion of a spelled word into a phonetic transcription and does not disclose two separate steps that include the searching of a dictionary. Finally, the combination fails to disclose the registering of the base form from the pronunciation dictionary in a speech recognition dictionary. Contolini contains no disclosure of a base form of a word or the registration of a base form from a pronunciation dictionary in a speech recognition dictionary.

The Examiner acknowledges that Hutchins and Contolini fail to disclose sounds-like spellings. While IBM TDB may make mention of the use of sounds-like spellings in building word models for use by a speech recognition system, lines 12-13 of IBM TDB state that the technique described therein “proposes that sounds-like spellings be given by the user when it is wished to provide better data to the spelling-to-sound rules.” Such rules may be used to add a word to the vocabulary of the speech recognition system.

obtain a sounds-like spelling corresponding to the word inscription. Thus, the user need not enter the sounds-like spelling since a word dictionary is searched to obtain a sounds-like spelling corresponding to said word inscription. A pronunciation dictionary is then searched to obtain a base form corresponding to the sounds-like spelling that has been obtained. Then, the base form is registered in a speech recognition dictionary. This is accomplished, for example as pointed out at page 13, line 21, of the present specification, “without a voice having to be recorded.” This is not what is disclosed by the IBM TDB technique, and certainly not by any combination of IBM TDB with Hutchins and Contolini.

Appellants do not assert that they have developed the concept of “sounds-like spellings.” Such spellings are known, for example, as evidenced by IBM TDB. However, the automated use of “sounds-like spellings” as recited in the registration techniques of the claimed invention was not known prior to the invention and is clearly not taught or suggested by the combination of Hutchins, Contolini and IBM TDB.

Thus, for at least the above reasons, Appellants request withdrawal of the §103(a) rejection of claims 1, 5 and 9.

Similar arguments apply to independent claims 2, 6 and 10, which also recite a method, apparatus and program code, respectively, for performing recognized word registration. In accordance with the claimed invention, a word inscription is specified by a user. A word dictionary is searched to obtain a plurality of sounds-like spellings that correspond to said word inscription and sounds-like spelling scores that correspond to the sounds-like spellings. The plurality of sounds-like spellings are displayed for the user. The sounds-like spelling that is selected by the user is obtained from among the plurality of sounds-like spellings. Then, a pronunciation dictionary is searched to obtain a base form and a pronunciation score corresponding to the sounds-like spelling that has been obtained. A determination is made as to whether the pronunciation score exceeds a predetermined threshold value. The base form is then registered in a speech recognition dictionary when the pronunciation score exceeds the predetermined threshold value.

For the same reasons as pointed out above, the combination of Hutchins, Contolini and IBM TDB do not teach searching a word dictionary to obtain a plurality of sounds-like spellings that

correspond to the word inscription and sounds-like spelling scores that correspond to the sounds-like spellings; displaying the plurality of sounds-like spellings for the user; obtaining the sounds-like spelling that is selected by the user from among the plurality of sounds-like spellings; searching a pronunciation dictionary to obtain a base form and a pronunciation score corresponding to the sounds-like spelling that has been obtained; determining whether the pronunciation score exceeds a predetermined threshold value; and registering the base form in a speech recognition dictionary when the pronunciation score exceeds the predetermined threshold value. The Office Action seems to point to IBM TDB for disclosing sounds-like spelling scores, and displaying sounds-like spellings to the user; however, Appellants find no such teachings in the reference. Also, since Hutchins does not relate to a word registration process, nothing therein teaches or suggests the elements of claims 2, 6 and 10. Contolini and IBM TDB fails to remedy these deficiencies.

Thus, for at least the above reasons, Appellants request withdrawal of the §103(a) rejection of claims 2, 6 and 10.

Lastly, similar arguments apply to independent claims 3, 7 and 11, which also recite a method, apparatus and program code, respectively, for performing recognized word registration. The claimed invention determines whether first voice information obtained for a user's voice matches a predetermined condition. A speech recognition wizard panel that includes a new word input field and a sounds-like spelling input field is displayed on a display screen, when the voice information matches said predetermined condition. A new word and a sounds-like spelling that are entered in the speech recognition wizard panel are obtained. Second voice information based on the user's pronunciation provided for the new word and the sounds-like spelling is obtained. The second voice information, the new word and the sounds-like spelling are employed to specifically describe a base form. Then, the base form is added to a speech recognition dictionary.

Hutchins, Contolini and IBM TDB fail to teach or suggest such a word registration technique that uses a speech recognition wizard panel, and such first and second voice information from the user. The Office Action seems to point to Hutchins for disclosing such a wizard panel and use of such user voice information; however, Hutchins does not relate to a word registration process. Also, no such details are disclosed in Contolini or IBM TDB.

Thus, for at least the above reasons, Appellants request withdrawal of the §103(a) rejection of claims 3, 7 and 11.

Appellants point out that while their arguments may, at times, discuss Hutchins, Contolini and IBM TDB one at a time, it is in an effort to clearly illustrate that the references do not teach or suggest one or more of the elements of the claimed invention. Thus, in general, by pointing out that reference 1 fails to disclose an element, say element A, and then pointing out that references 2 and 3 also fails to disclose element A, the discussion effectively points out that the combination of the three references, even if proper, would fail to disclose element A.

For at least the reasons given above, Appellants respectfully request withdrawal of the §103(a) rejections of claims 1-3, 5-7 and 9-11. Appellants believe that claims 1-3, 5-7 and 9-11 are not obvious in view of Hutchins, Contolini and IBM TDB. As such, the application is asserted to be in condition for allowance, and favorable action is respectfully solicited.

Respectfully submitted,

A handwritten signature in black ink, appearing to read "Robert W. Griffith". The signature is fluid and cursive, with the first name "Robert" and last name "Griffith" being clearly legible despite the cursive style.

Date: November 26, 2003

Robert W. Griffith
Attorney for Applicant(s)
Reg. No. 48,956
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-4547

APPENDIX

1. A recognized word registration method, for a speech recognition apparatus that includes a display screen and a voice input device, comprising the steps of:

obtaining a word inscription specified by a user;

searching a word dictionary to obtain a sounds-like spelling corresponding to said word inscription;

searching a pronunciation dictionary to obtain a base form corresponding to said sounds-like spelling that has been obtained; and

registering said base form in a speech recognition dictionary.

2. A recognized word registration method, for a speech recognition apparatus that includes a display screen and a voice input device, comprising the steps of:

obtaining a word inscription specified by a user;

searching a word dictionary to obtain a plurality of sounds-like spellings that correspond to said word inscription and sounds-like spelling scores that correspond to said sounds-like spellings;

displaying said plurality of sounds-like spellings for said user;

obtaining said sounds-like spelling that is selected by said user from among said plurality of sounds-like spellings;

searching a pronunciation dictionary to obtain a base form and a pronunciation score corresponding to said sounds-like spelling that has been obtained;

determining whether said pronunciation score exceeds a predetermined threshold value; and

registering said base form in a speech recognition dictionary when said pronunciation score exceeds said predetermined threshold value.

3. A recognized word registration method, for a speech recognition apparatus that includes a display screen and a voice input device, comprising the steps of:

determining whether first voice information obtained for a user's voice matches a predetermined condition;

displaying on said display screen, when said voice information matches said predetermined condition, a speech recognition wizard panel that includes a new word input field and a sounds-like spelling input field;

obtaining a new word and a sounds-like spelling that are entered in said speech recognition wizard panel;

obtaining second voice information based on said user's pronunciation provided for said new word and said sounds-like spelling;

employing said second voice information, said new word and said sounds-like spelling to specifically describe a base form; and

adding said base form to a speech recognition dictionary.

5. A speech recognition apparatus, which includes a display screen and a voice input device, comprising:

a recognized word registration unit for obtaining a word inscription specified by a user;

a sounds-like spelling generator for searching a word dictionary to obtain a sounds-like spelling corresponding to said word inscription;

a base form generator for searching a pronunciation dictionary to obtain a base form corresponding to said sounds-like spelling that has been obtained; and

a speech recognition dictionary in which said base form is registered.

6. A speech recognition apparatus, which includes a display screen and a voice input device, comprising:

a recognized word registration unit for obtaining a word inscription specified by a user;

a sounds-like spelling generator for searching a word dictionary to obtain a plurality of sounds-like spellings that correspond to said word inscription and sounds-like spelling scores that correspond to said sounds-like spellings, and for obtaining said sounds-like spelling that is selected by said user from among said plurality of sounds-like spellings on said display screen;

a base form generator for searching a pronunciation dictionary to obtain a base form and a pronunciation score corresponding to said sounds-like spelling that has been obtained; and

a speech recognition dictionary in which said base form is registered when said pronunciation score exceeds a predetermined threshold value.

7. A speech recognition apparatus, which includes a display screen and a voice input device, comprising:

a recognized word registration unit for determining whether first voice information obtained for a user's voice matches a predetermined condition;

a speech recognition wizard for displaying on said display screen, when said voice information matches said predetermined condition, a speech recognition wizard panel that includes a new word input field and a sounds-like spelling input field;

a voice input unit for obtaining second voice information based on said user's pronunciation provided for a new word and a sounds-like spelling that are entered in said speech recognition wizard panel;

a base form generator for employing said second voice information, said new word and said sounds-like spelling to specifically describe a base form; and

a speech recognition dictionary to which said base form is added.

9. A storage medium on which is stored a recognized word registration program that is to be executed by a speech recognition apparatus that includes a display screen and a voice input device, said recognized word registration program comprising:

program code for instructing said speech recognition apparatus to obtain a word inscription specified by a user;

program code for instructing said speech recognition apparatus to search a word dictionary to obtain a sounds-like spelling corresponding to said word inscription;

program code for instructing said speech recognition apparatus to search a pronunciation dictionary to obtain a base form corresponding to said sounds-like spelling that has been obtained; and

program code for instructing said speech recognition apparatus to register said base form in a speech recognition dictionary.

10. A storage medium on which is stored a recognized word registration program that is to be executed by a speech recognition apparatus that includes a display screen and a voice input device, said recognized word registration program comprising:

program code for instructing said speech recognition apparatus to obtain a word inscription specified by a user;

program code for instructing said speech recognition apparatus to search a word dictionary to obtain a plurality of sounds-like spellings that correspond to said word inscription and sounds-like spelling scores that correspond to said sounds-like spellings;

program code for instructing said speech recognition apparatus to display said plurality of sounds-like spellings for said user;

program code for instructing said speech recognition apparatus to obtain said sounds-like spelling that is selected by said user from among said plurality of sounds-like spellings;

program code for instructing said speech recognition apparatus to search a pronunciation dictionary to obtain a base form and a pronunciation score corresponding to said sounds-like spelling that has been obtained;

program code for instructing said speech recognition apparatus to determine whether said pronunciation score exceeds a predetermined threshold value; and

program code for instructing said speech recognition apparatus to register said base form in a speech recognition dictionary when said pronunciation score exceeds said predetermined threshold value.

11. A storage medium on which is stored a speech recognition process program that is to be executed by a speech recognition apparatus that includes a display screen and a voice input device, said speech recognition process program comprising:

program code for instructing said speech recognition apparatus to determine whether first voice information obtained for a user's voice matches a predetermined condition;

program code for instructing said speech recognition apparatus to display on said display screen, when said voice information matches said predetermined condition, a speech recognition wizard panel that includes a new word input field and a sounds-like spelling input field;

program code for instructing said speech recognition apparatus to obtain a new word and a sounds-like spelling that are entered in said speech recognition wizard panel;

program code for instructing said speech recognition apparatus to obtain second voice information based on said user's pronunciation provided for said new word and said sounds-like spelling;

program code for instructing said speech recognition apparatus to employ said second voice information, said new word and said sounds-like spelling to specifically describe a base form; and

program code for instructing said speech recognition apparatus to add said base form to a speech recognition dictionary.